

Secure Multi-Source Multicast

Alejandro Cohen
BGU

Asaf Cohen
BGU

Muriel Médard
MIT

Omer Gurewitz
BGU

Abstract—The principal mission of *Multi-Source Multicast* (MSM) is to disseminate all messages from all sources in a network to all destinations. MSM is utilized in numerous applications. In many of them, securing the messages disseminated is critical.

A common secure model is to consider a network where there is an eavesdropper which is able to observe a subset of the network links, and seek a code which keeps the eavesdropper ignorant regarding *all the messages*. While this is solved when all messages are located at a single source, *Secure MSM* (SMSM) is an open problem, and the rates required are hard to characterize in general.

In this paper, we consider *Individual Security*, which promises that the eavesdropper has zero mutual information with *each message individually*. We completely characterize the rate region for SMSM under individual security, and show that such a security level is achievable at the full capacity of the network, that is, the cut-set bound is the matching converse, similar to *non-secure* MSM. Moreover, we show that the field size is similar to non-secure MSM and does not have to be larger due to the security constraint.

I. INTRODUCTION

Linear Network Coding (LNC) [1] and Random Linear Network Coding (RLNC) [2] are essential for efficient utilization of network resources. With network coding, *multiple sources* can multicast information to all destinations simultaneously, at rates up to the min-cut between the sources and the destinations. Figure 1 depicts a simple example: the min-cut from any source to any destination is 2, and from both sources to any destination is 4, one can disseminate *2 messages from each source to all destinations*. However, in many practical multicast applications, it is important to assure privacy is not compromised if an eavesdropper (Eve) is present in the network. Indeed, the theory of secure network coding is vast. We include here only the most relevant works.

When the sources are co-located at a single node, several secure network coding solutions were suggested [3]–[5]. Such solutions guarantee the mutual information between Eve’s data, \mathbf{Z} , and all the messages is 0. For example, returning to Figure 1, if only source s_1 had messages to send, and Eve would be able to wiretap one link in the network, then secure network coding would guarantee secure dissemination of one message from the source to all destinations. This is a reduction in rate compared to the full capacity, as the min-cut from s_1 to any destination is 2.

When the network includes multiple sources which are not co-located, the problem is more involved. Clearly, applying a single-source, secure network coding solution at each source would give an achievable scheme. In the example, if Eve wiretaps one link, one can clearly multicast one message *from each source, to all destinations*. This solution, however, may be wasteful, as it is half of the full capacity of the network, “wasting” one message *per source*, although Eve may capture

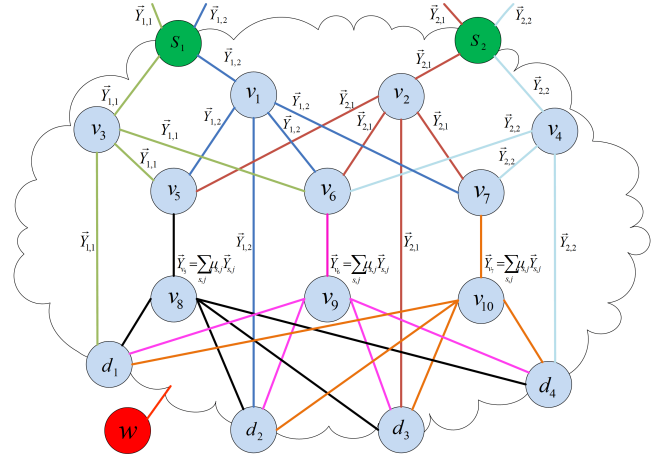


Figure 1: Secure multi-source multicast with LNC, for two sources s_i , with two messages each and four legitimate destination nodes d_i . The eavesdropper min-cut is at most 1. The edges in the graph point downward.

only a single link regardless of the number of sources. Indeed, there is no matching converse result for the above solution.

In [6], [7], the authors gave a necessary and sufficient condition for Secure Multi-Source Multicast (SMSM). However, it is a condition on *ranks of matrices* having the global encoding vectors as columns, and, unlike non-secure MSM or secure single-source multicast, it does not translate directly to *rate or min-cut constraints*. In [8], the authors characterized the network coding capacity of several models, including SMSM, via the entropic region Γ^* . Yet, to date, this region is not fully characterized. Thus, the problem of determining the rate region in SMSM remains an open problem in general [9]. As mentioned in [10, Section VI], seeking models for which it is solvable is important.

Main Contribution

In this paper, we consider SMSM under an *Individual Security* constraint. In this model, the eavesdropper is kept ignorant, in the sense of having zero mutual information, regarding each message separately, yet may potentially obtain *insignificant* information about mixtures of packets transmitted. Such a security model was recently used in various problems, e.g., wiretap channels [11], more general broadcast channels [12]–[15], multiple-access channels [16], [17], multiple-input multiple-output channels [18], and is also related to notions of *algebraic security* [19], [20] which consider the information in linear combinations of messages.

We completely characterize the rate region for individually secure MSM. Specifically, we show that secure communication is achievable up to the min-cut, that is, without any decrease in the rate or any message “blow-up” by extra randomness. In

fact, due to the individual security constraint, messages protect one another, and in the context of Figure 1, one is able to send *two messages from each source securely, although Eve may observe any single link*. Finally, we briefly show that the coding scheme is applicable to algebraic gossip as well [21], resulting in *secure gossip* without extra rounds.

II. MODEL AND PROBLEM FORMULATION

SMSM is specified by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} and \mathcal{E} are the node set and the edge set, respectively. We assume noise-free links of unit capacity. This capacity can be thought of as one "packet" of c bits, plus some negligible overhead.¹

The node set \mathcal{V} contains a subset of source nodes $\mathcal{S} = \{S_1, \dots, S_{|\mathcal{S}|}\}$ and a subset of legitimate destination nodes $\mathcal{D} = \{D_1, \dots, D_{|\mathcal{D}|}\}$. Each of the sources has its own set of k independent messages of length c each, over the binary field. We denote them by a messages matrix

$$\mathbf{M}_s = [\vec{M}_{s,1}; \vec{M}_{s,2}; \dots; \vec{M}_{s,k}] \in \{0, 1\}^{k \times c},$$

where each row corresponds to a separate message $\vec{M}_{s,j}$, $j \in \{1, \dots, k\}$.

We assume an eavesdropper which can obtain a subset of w packets traversing the network. Specifically, we define the eavesdropper matrix as

$$\mathbf{Z}_w = [Z_1^c; Z_2^c; \dots; Z_w^c] \in \{0, 1\}^{w \times c}.$$

For $s_i \in \mathcal{S}$ and $d_i \in \mathcal{D}$, we denote the values of min-cuts in the network by $\rho(\cdot; \cdot)$. For example, $\rho(s_1; d_1)$ represents the value of the min-cut from source node s_1 to legitimate node d_1 . $\rho(s_1; z)$ represents the value of the min-cut from source node s_1 to the eavesdropper (assuming z is a virtual node with infinite capacity from the w edges observed by Eve) and $\rho(\mathcal{S}; d_1)$ represents the value of the min-cut from all the source nodes to legitimate node d_1 .

The goal is to design secure multi-source multicast coding where legitimate nodes send their available messages in order to disseminate all the messages to all the legitimate destination nodes, yet, observing w packets from the communication between legitimate nodes, the eavesdropper is ignorant regarding the messages.

Definition 1. An MSM algorithm with parameters k and w is *Reliable* and *Individually* or *Strongly* secure if:

(1) *Reliable*: At the legitimate destination node $d \in \mathcal{D}$, letting \mathbf{Y}_d denote the message matrix obtained, for any set of messages \mathbf{M}_s , $s \in \mathcal{S}$, we have

$$P(\hat{\mathbf{M}}_s(\mathbf{Y}_d) \neq \mathbf{M}_s) \leq \epsilon,$$

where $\hat{\mathbf{M}}_s(\mathbf{Y}_d)$ is the estimation of messages \mathbf{M}_s at d .

(2) *Individually* secure: At the eavesdropper, observing w packets, we have

$$H(M_{s,j} | \mathbf{Z}_w) = H(M_{s,j}),$$

for all $j \in \{1, \dots, k\}$ and for all $s \in \mathcal{S}$.

¹As in most LNC solutions, a header is required for each message. Thus, we assume messages of length c , large enough to make the overhead in the header negligible.

(3) *Strongly* secure: At the eavesdropper, observing w packets, we have

$$H(\{\mathbf{M}_s\}_{s \in \mathcal{S}} | \mathbf{Z}_w) = H(\{\mathbf{M}_s\}_{s \in \mathcal{S}}).$$

Remark 1. The individual-secrecy constraint given in Definition 1.2 does not promise perfect, strong-secrecy [3], [5], [6], which is, having the mutual information with all messages negligible. Individual-secrecy ensures secrecy only on each message $M_{s,j}$ separately. The eavesdropper, observing \mathbf{Z}_w , may obtain some information on the combination of k messages since the messages are not independent given \mathbf{Z}_w . However, since the k original messages are mutually independent, the leaked information has no meaning [11]–[18]. In other words, since

$$I(\mathbf{M}_s; \mathbf{Z}_w) = \sum_k I(M_{s,k}; \mathbf{Z}_w | M_s^{k-1}) \geq \sum_k I(M_{s,k}; \mathbf{Z}_w),$$

we require that the r.h.s will be small, however, this does not guarantee that the l.h.s is small. If the eavesdropper receives message $M_{s,j}$ by any other manner than the Individual-SMSM transmissions, Eve may obtain some information on other messages $M_{s,i}$, $i \neq j$, from $M_{s,j}$ and \mathbf{Z}_w . If it is required to prevent the possibility of such an attack, one can get perfect secrecy using Definition 1.3, yet at the price of a higher rate as given in Appendix A.

Remark 2. For multicast problems and LNC, the condition in (1) can be used with $\epsilon = 0$ [1], [2]. Yet, we allow a small error to cope with protocols such as gossip [21], [22].

A. Source and Network Coding

We assume a source $s \in \mathcal{S}$ may use an encoder,

$$f : \mathcal{M}_s \rightarrow \mathcal{X}_s \in \{0, 1\}^{n \times c},$$

which maps each message matrix \mathbf{M}_s to a matrix \mathbf{X}_s of codewords. When using strong security constraints, e.g., [3], [5], $n > k$ and represents a message "blow-up" using a random key, used to confuse Eve. However, the main contribution herein, is that *under individual-secrecy*, $n = k$ suffices, and *there will be no rate loss due to the secrecy constraint*.

Then, the source packets \vec{Y} transmitted are linear combinations of $\{\vec{X}_r\}_{r=1}^n$ with coefficients in the usual LNC sense, i.e.,

$$\vec{Y} = \sum_{r=1}^n \mu_r \vec{X}_r.$$

Each node maintains a subspace Y_v that is the span of all packets known to it. In RLNC, when node v sends a packet, $Out(\vec{Y})$, it chooses uniformly a packet from Y_v by taking a random linear combination. If a deterministic algorithm is used, e.g., [23], the coefficients are calculated based on the network topology. The code we suggest herein is only at the sources, and then utilizes any capacity-achieving, non-secure network code.

B. Gossip in Oblivious Networks

While the results in this paper are tailored to LNC in the sense of [1], [2], they easily apply to *algebraic gossip* [21] as well. We briefly describe this model.

The network operates in rounds. In each round t , the sources, as well as any legitimate node which has messages it previously received, pick a random node to exchange information with. The information exchange is done by either sending (PUSH) or receiving (PULL) a message. In algebraic gossip, the message sent by a node v is simply a random linear combination of the vectors which form a basis for Y_v . The process stops when all the legitimate nodes have all the messages, i.e., have a full rank matrix. We briefly review the definitions and results from [22] for non-secure gossip networks, which we will use to formulate our result in this context.

Definition 2. A network is *oblivious* if the topology G_t at time t only depends on t , $G_{t'}$ for any $t' < t$ and some randomness. We call an oblivious network model furthermore i.i.d. if the topology G_t is independent of t and prior topologies.

Consider a single (uncoded) message, and the set of nodes S_l which received that message after l rounds. S_l advances like a flooding process F . We say that F stops at time t if the message is received at all nodes after t rounds. Let S_F be the random variable denoting the stopping time of F .

Definition 3. We say an oblivious network with a vertex set V floods in time T with throughput α if there exists a prime power q such that for every vertex $v \in V$ and every $k > 0$ we have $P[S_F \geq T + k] < q^{-\alpha k}$.

III. MAIN RESULTS

The main result is the following achievability theorem, which states that individually-secure multi-source multicast is achievable at rates up to the network min-cuts using LNC.

A. Individually Secure MSM

Theorem 1. Assume an SMSM network $(\mathcal{V}, \mathcal{E}, \mathcal{S}, \mathcal{D}, w)$. There exists a coding scheme which disseminates k messages from each source in \mathcal{S} , to all destinations in \mathcal{D} , while keeping an eavesdropper which observes w links ignorant of with respect to each message individually if:

- 1) For all $s \in \mathcal{S}$ and all $d \in \mathcal{D}$, $\rho(s, d) \geq k$.
- 2) For all $d \in \mathcal{D}$, $\rho(\mathcal{S}, d) \geq k|\mathcal{S}|$.

Under strong-secrecy, i.e., requiring Eve's mutual information with all messages simultaneously to be zero, the problem of MSM is still open [9], [10, Section VI]. Clearly, if Eve observes w links, a naive implementation, which increases the message rates from each source by w , can send k messages from each source and achieve strong secrecy if $\rho(s, d) \geq k+w$ and $\rho(\mathcal{S}, d) \geq (k+w)|\mathcal{S}|$ for each $s \in \mathcal{S}$ and $d \in \mathcal{D}$. However, such an implementation is clearly wasteful, and, to date, the optimal strategy is unknown. Obviously, the required rates under strong secrecy are higher than the min-cut bound, as even for single-source multicast one needs $\rho(s_1, d_i) \geq k+w$ [3]. In Appendix A, we provide a code for Strong-SMSM. It is important to note that in the code suggested, the alphabet size does not increase with the network parameters due to the strong-security constraint.

The importance of Theorem 1 is that under individual secrecy, not only the rate region can be characterized, and is achievable using linear network coding, individually secure MSM is possible up to the min-cuts in the network.

The tightness of Theorem 1, in terms of rates, is trivially achieved using the cut-set bound. That is, the conditions $\rho(s, d) \geq k$ and $\rho(\mathcal{S}, d) \geq k|\mathcal{S}|$ are required solely to achieve reliability, nevertheless when security is an additional constraint. However, a stronger notion of a converse can be given. To this end, we first note that Theorem 1 guarantees a slightly weaker level of security than is possible at the same rates. Specifically, Theorem 1 guarantees $I(M_{s,j}; \mathbf{Z}) = 0$ for any single message j . However, ensuring the mapping from \mathbf{M}_s to \mathbf{X}_s mixes the messages appropriately, i.e., satisfies rank constraints similar to [10, Lemma 3.1], can, in fact, ensure Eve is kept ignorant of any set of $k-w$ messages. That is, guarantee a “set”-individual secrecy with respect to any set of $k-w$ messages. This is also consistent with the secrecy guarantees in [11]. Under such an individual secrecy constraint, the converse below gives a stronger result than the min-cut bound.

Theorem 2. Assume an SMSM network $(\mathcal{V}, \mathcal{E}, \mathcal{S}, \mathcal{D}, w)$. Under individual security for $k-w$ messages, that is, requiring $I(\mathbf{M}_s^{k-w}; \mathbf{Z}_w) = 0$ for any set of $k-w$ messages, one must have

$$H(\mathbf{M}_s) \leq \rho(s, d_i) - \rho(s, z) + w.$$

This result should be interpreted as follows. If Eve observes w independent links, and $\rho(s; z) = w$, then one must have $H(\mathbf{M}_s) \leq \rho(s, d_i)$, which is the cut set bound. However, if Eve observes more than w links, and one wishes to maintain the individual-secrecy constraint, then $H(\mathbf{M}_s)$ should be strictly smaller than $\rho(s, d_i)$ and at the same amount. E.g., if Eve observes $w + e$ links, we have $H(\mathbf{M}_s) \leq \rho(s, d_i) - e$.

As mentioned earlier, the suggested code easily applies to algebraic gossip as well, since this can be viewed as linear network coding over a time-extended graph. The following results capture the number of rounds required to (individually) securely disseminate k messages from each of the $|\mathcal{S}|$ sources to all nodes in the network.

Theorem 3. Assume an oblivious network that floods in time T with throughput α . Then, for $|\mathcal{S}|$ nodes in the network with k messages each, algebraic gossip spreads the $k|\mathcal{S}|$ messages to all nodes with probability $1 - \epsilon$ after

$$T' = T + \frac{1}{\alpha}(k|\mathcal{S}| + \log \epsilon^{-1})$$

rounds, while keeping any eavesdropper which observes at most w packets ignorant about each message individually.

Thus, compared to only a reliability constraint, the number of rounds required for both reliability and individual-secrecy is exactly the same as in the original non-secure gossip protocol.

B. Alphabet Size

Without secrecy constraints, Jaggi *et al.* proved that a field with size greater than or equal to the number of destinations is sufficient to multicast LNC [23]. However, this may not hold if it is required to keep an eavesdropper ignorant. Cai *et al.*

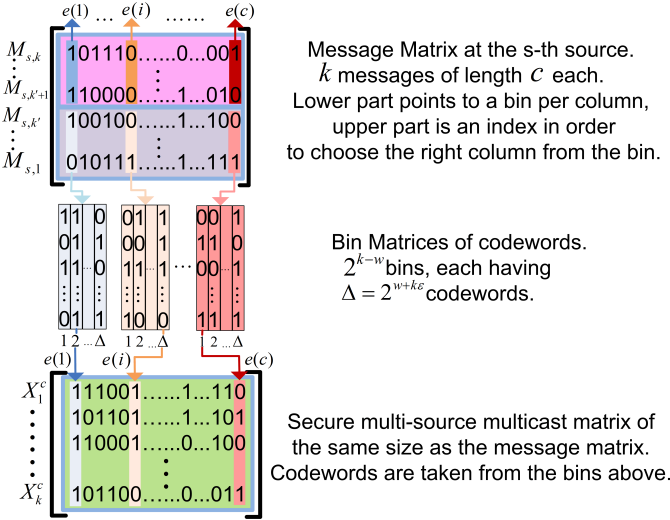


Figure 2: Binning and source encoding process for Individual-SMSM.

[3] devised a code which requires a field of exponential size to obtain secrecy. There, the field size must be larger than $\binom{|E|}{w}$. Feldman *et al.* [24] showed that there exist networks that require a field of size at least $\Theta(|E|^{\frac{w}{2}})$. In [5], the authors considered the canonical wiretap II as a special case of the wiretap network II . When d is the number of destinations in the multicast connection, a field of size $\binom{2k^3 d^2}{w-1+d}$ is sufficient, which is independent of $|V|$ and $|E|$ but still exponential in other network parameters.

In the solution we suggest herein, the field size is determined only by the network coding scheme, that is, only by the requirement for *reliability*, and is not increased by the individual-security constraints. In the gossip case, for example, since $q^{-\alpha k} = 2^{-(\alpha \log q)k}$, any field size greater than or equal to 2 will suffice, and α decrease in the field size has only a logarithmic effect on the throughput, meaning only a logarithmic multiplier on the number of rounds T' required.

IV. CODE CONSTRUCTION AND A PROOF FOR INDIVIDUAL-SMSM (THEOREM 1 AND THEOREM 3)

At each source node $s \in \{1, \dots, |S|\}$, we map each column of the message matrix \mathbf{M}_s . Specifically, as depicted in Figure 2, in the code construction phase, for each partial column $M_{s,1}(i); \dots; M_{s,k'}(i)$ of length k' in the message matrix we generate a bin, containing several columns of length k . The number of such columns *corresponds* to w , the number of packets that the eavesdropper can wiretap, in a relation that will be made formal in the sequel. Then, for the i -th column of the s -th message matrix, the selected column from the i -th bin therein is according to $M_{s,k'+1}(i); \dots; M_{s,k}(i)$. That is, the upper part of the original column serves an *index* in order to choose the right column from the bin. This way, a new, $k \times c$ message matrix \mathbf{X}_s is created. This message matrix contains k new messages of the same size c . We may now turn to the detailed construction and analysis.

1) *Codebook Generation*: Set $\Delta = 2^{w+k\epsilon}$. Let $P(x) \sim \text{Bernoulli}(1/2)$. Using a distribution $P(X^k) = \prod_{j=1}^k P(x_j)$,

for each possible column $M_1(i); \dots; M_{k'}(i)$ in the message matrix, generate Δ independent and identically distributed codewords $x^k(e)$, $1 \leq e \leq \Delta$, where ϵ can be chosen arbitrarily small. Note the codebook matrix is of the same size as \mathbf{M}_s . The codebook is depicted in Figure 2.

2) *Source and legitimate Node encodings*: At each s -th source node, the encoder selects, for each column i of $M_{s,1}(i); \dots; M_{s,k'}(i)$ bits, one codeword $x^k(e(i))$ from the i -th bin, where $e(i) = M_{s,k'+1}(i); \dots; M_{s,k}(i)$. That is, $k' = k - w$ bits of the column choose the bin, and the remaining w bits choose the codeword within the bin. Then, the sources transmit linear combinations of the rows, with random coefficients. Nodes transmit random linear combinations of the vectors in \mathcal{S}_v , which is maintained by each node according to the messages received at the node.

A. Reliability

The reliability proof using RLNC is almost a direct consequence of [2]. Clearly, the min-cut is given by Theorem 1. Hence, the legitimate nodes can easily reconstruct \mathbf{X}_s for each s (simple, non-secure, multi-source multicast). Then, each destination maps \mathbf{X}_s back to \mathbf{M}_s , as this is a 1 : 1 mapping.

In the same way, using a gossip protocol, the reliability proof is almost a direct consequence of [22, Theorem 1]. Hence, the number of rounds required is given by Theorem 3.

An example, obtaining both reliability and individual secrecy for two sources, with two messages each and four legitimate destination nodes, where the eavesdropper min-cut is at most 1, is given in Figure 1. Note that secure communication with respect to *one message* is possible while sending two messages from each source to all destinations.

Remark 3. In the code suggested herein, there is no required asymptotic for the size of k or c , which is usually required in information theoretic security solutions. Individual-secrecy in SMSM networks holds for any k satisfying

$$k \geq \lceil \rho(s, d) / (\rho(s, d) - \rho(s; z)) \rceil \geq 2.$$

B. Information Leakage at the Eavesdropper

We now prove the individual-security constraint is met, using techniques recently given in [25]. In particular, for the individual constraint, we wish to show that $I(M_{s,j}; \mathbf{Z}_w)$ is small for all $s \in \mathcal{S}$ and all j . We will do that by showing that given \mathbf{Z}_w , Eve's information, all possibilities for $M_{s,j}$ are equally likely, hence Eve has no intelligent estimation for $M_{s,j}$.

Denote by \mathcal{C}_k the random codebook and by \mathbf{X}_s the set of codewords corresponding to $\tilde{M}_{s,1} \dots \tilde{M}_{s,k}$. To analyze the information leakage at the eavesdropper, note that Eve has access to at most w linear combinations on the rows of \mathbf{X}_s . We will assume these linear combinations are in fact independent, and since Eve has access to the coefficients, we will assume Eve can even use Gaussian elimination and have access to w rows from original matrix \mathbf{X}_s .

Next, note that the columns of \mathbf{X}_s are independent (by the construction of the codebook, creating \mathbf{X}_s is done independently per-column; c columns are used only to reduce the

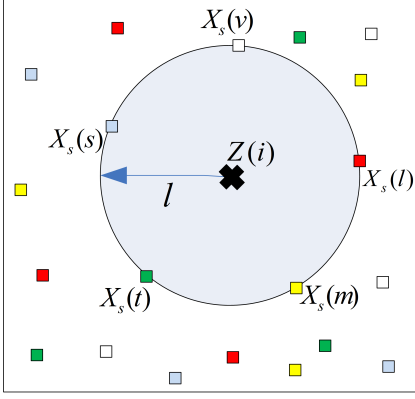


Figure 3: Codewords for Individual-SMSM algorithm lie exactly in a ball of radius $l = k - w$ around Z .

NC overhead). Hence, it suffices to consider the information leakage for each column $i \in \{1, \dots, c\}$ from \mathbf{X}_s separately.

For each column i of \mathbf{M}_s , the encoder has Δ independent and identically distributed codewords, out of which one is selected. Hence, there is an exponential number of codewords, from the eavesdropper's perspective, that can generate a column in \mathbf{X}_s , and we require that Eve is still confused even given the w bit from each column. Let $\mathbf{Z}_w(i)$ be the w bits Eve has from column i . Denote $l = k - w$. Define by $Sh(\mathbf{Z}_w(i), l)$ the set of all k -tuples consistent with $\mathbf{Z}_w(i)$, i.e.,

$$Sh(\mathbf{Z}_w(i), l) = \{b^k : b^k(\mathcal{S}_Z) = \mathbf{Z}_w(i)\},$$

where \mathcal{S}_Z denotes indices of the rows Eve has. Clearly, there are 2^l tuples in $Sh(\mathbf{Z}_w(i), l)$. See Figure 3 for a graphical illustration. We assume Eve has the codebook, yet does not know which column from each bin is selected to be the codeword. Hence, we wish to show that given $\mathbf{Z}_w(i)$, Eve will have at least one candidate per bin. The probability for a codeword to fall in a given shell is

$$\begin{aligned} Pr(\mathbf{X}_s^k(i) \in \mathcal{C}_k \cap \mathbf{X}_s^k(i) \in Sh(\mathbf{Z}_w(i), l)) \\ = \frac{Vol(Sh(\mathbf{Z}_w(i), l))}{2^k} = \frac{2^{(k-w)}}{2^k}. \end{aligned}$$

In each bin of \mathcal{C}_k , we have $\Delta = 2^{w+k\epsilon}$ codewords. Thus, the number of codewords Eve sees on a shell, *per bin* is

$$|\{m(i) : X^k(i) \in Sh(Z(i), l)\}| = \frac{2^{w+k\epsilon} * 2^{k-w}}{2^k} = 2^{k\epsilon}.$$

Hence, we can conclude that on average, and if $k\epsilon$ is not too small, for every column in \mathbf{M}_s Eve has a few possibilities in each bin, hence cannot locate the right bin. However, it is still important to show that all bins have (asymptotically) equally likely number of candidate codewords, hence Eve cannot locate a preferred bin.

To this end, we proved that the average number of codewords per column is $2^{k\epsilon}$. We wish to show that now the probability that the actual number of options deviates from the average by more than ϵ is small. Define

$$\begin{aligned} \mathcal{E}_{C_1}(Z(i), l) &:= Pr\{(1 - \epsilon)2^{k\epsilon} \leq \\ &|m(i) : X_s^k(i) \in Sh(\mathbf{Z}_w(i), l)| \leq (1 + \epsilon)2^{k\epsilon}\}. \end{aligned}$$

By the Chernoff bound, we have

$$Pr(\mathcal{E}_{C_1}(Z(i), l)) \geq 1 - 2^{-\epsilon' 2^{k\epsilon}}.$$

Due to the super exponential decay in k , when taking a union bound over all columns, the probability that Eve decodes correctly some column is small. Hence, for Eve, all codewords are almost equiprobable and $I(M_{s,j}; \mathbf{Z}_w) \rightarrow 0$.

V. CONVERSE (THEOREM 2)

In this section, we derive a converse result, which shows that under individual secrecy on a group of $k - w$ messages, not only the rate is bounded by the min-cut, but, more importantly, any independent link that Eve observes above w will require to reduce the rate *at the same amount* in order to achieve both reliability and secrecy.

Let $\bar{\mathbf{Z}}$ denote the random variable corresponding to the links which are not available to Eve. Hence, $\mathbf{Y}_d = (\mathbf{Z}, \bar{\mathbf{Z}})$. Let \mathbf{M}_s^{k-w} denote a set of $k - w$ messages, and \mathbf{M}_s^w denote the remaining w . We will show that reliability, that is $H(\mathbf{M}_s | \mathbf{Y}_d) = 0$, and individual secrecy, that is, $I(\mathbf{M}_s^{k-w}; \mathbf{Z}) = 0$, imply that $H(\mathbf{M}_s)$ is upper bounded by the term in Theorem 2.

$$\begin{aligned} H(\mathbf{M}_s) &= H(\mathbf{M}_s^{k-w} | \mathbf{M}_s^w) + H(\mathbf{M}_s^w) \\ &\stackrel{(a)}{\leq} I(\mathbf{M}_s^{k-w}; \mathbf{Y}_d | \mathbf{M}_s^w) + H(\mathbf{M}_s^{k-w} | \mathbf{Y}_d) + w \\ &\stackrel{(b)}{\leq} I(\mathbf{M}_s^{k-w}; \mathbf{Z}, \bar{\mathbf{Z}} | \mathbf{M}_s^w) + w \\ &= I(\mathbf{M}_s^{k-w}; \mathbf{Z} | \mathbf{M}_s^w) + I(\mathbf{M}_s^{k-w}; \bar{\mathbf{Z}} | \mathbf{Z}, \mathbf{M}_s^w) + w \\ &= I(\mathbf{M}_s^{k-w}; \mathbf{Z}) + I(\mathbf{M}_s^w; \mathbf{Z} | \mathbf{M}_s^{k-w}) - I(\mathbf{Z}; \mathbf{M}_s^w) \\ &\quad + I(\mathbf{M}_s^{k-w}; \bar{\mathbf{Z}} | \mathbf{Z}, \mathbf{M}_s^w) + w \\ &\stackrel{(c)}{=} I(\mathbf{M}_s^w; \mathbf{Z} | \mathbf{M}_s^{k-w}) - I(\mathbf{Z}; \mathbf{M}_s^w) + I(\mathbf{M}_s^{k-w}; \bar{\mathbf{Z}} | \mathbf{Z}, \mathbf{M}_s^w) + w \\ &= I(\mathbf{M}_s^w; \mathbf{Z} | \mathbf{M}_s^{k-w}) - I(\mathbf{Z}; \mathbf{M}_s^w) \\ &\quad + H(\bar{\mathbf{Z}} | \mathbf{Z}, \mathbf{M}_s^w) - H(\bar{\mathbf{Z}} | \mathbf{M}_s^{k-w}, \mathbf{Z}, \mathbf{M}_s^w) + w \\ &= I(\mathbf{M}_s^w; \mathbf{Z} | \mathbf{M}_s^{k-w}) - I(\mathbf{Z}; \mathbf{M}_s^w) + H(\bar{\mathbf{Z}} | \mathbf{Z}, \mathbf{M}_s^w) + w \\ &= H(\mathbf{M}_s^w | \mathbf{M}_s^{k-w}) - H(\mathbf{M}_s^w | \mathbf{Z}, \mathbf{M}_s^{k-w}) - H(\mathbf{M}_s^w) + H(\mathbf{M}_s^w | \mathbf{Z}) \\ &\quad + H(\bar{\mathbf{Z}} | \mathbf{Z}, \mathbf{M}_s^w) + H(\bar{\mathbf{Z}}) - H(\bar{\mathbf{Z}}) + w \\ &= I(\mathbf{M}_s^w; \mathbf{M}_s^{k-w} | \mathbf{Z}) - I(\bar{\mathbf{Z}}; \mathbf{Z}, \mathbf{M}_s^w) + H(\bar{\mathbf{Z}}) + w \\ &= I(\mathbf{M}_s^w; \mathbf{M}_s^{k-w} | \mathbf{Z}) - I(\bar{\mathbf{Z}}; \mathbf{M}_s^w | \mathbf{Z}) - I(\bar{\mathbf{Z}}; \mathbf{Z}) + H(\bar{\mathbf{Z}}) + w \\ &\leq I(\mathbf{M}_s^w; \mathbf{M}_s^{k-w} | \mathbf{Z}) - I(\bar{\mathbf{Z}}; \mathbf{M}_s^w | \mathbf{Z}) + H(\bar{\mathbf{Z}}) + w \\ &= H(\mathbf{M}_s^w | \mathbf{Z}) - H(\mathbf{M}_s^w | \mathbf{Z}, \mathbf{M}_s^{k-w}) - H(\mathbf{M}_s^w | \mathbf{Z}) \\ &\quad + H(\mathbf{M}_s^w | \mathbf{Z}, \bar{\mathbf{Z}}) + H(\bar{\mathbf{Z}}) + w \\ &\leq H(\bar{\mathbf{Z}}) + w \\ &\stackrel{(d)}{\leq} \rho(s_i; d_i) - \rho(s_i; z) + w, \end{aligned}$$

where (a) is since conditioning reduces entropy, (b) is due to the reliability constraint, (c) follows since we assume that Eve is kept ignorant regarding any group of $w - k$ messages, hence $I(\mathbf{M}_s^{k-w}; \mathbf{Z}) = 0$, and (d) follows since $\rho(s_i; d_i) - \rho(s_i; z)$ is the maximum amount that may not be available to Eve, if she has a min-cut $\rho(s_i; z)$. Again, we assume unit capacity links and normalize the information in a message to "1" accordingly.

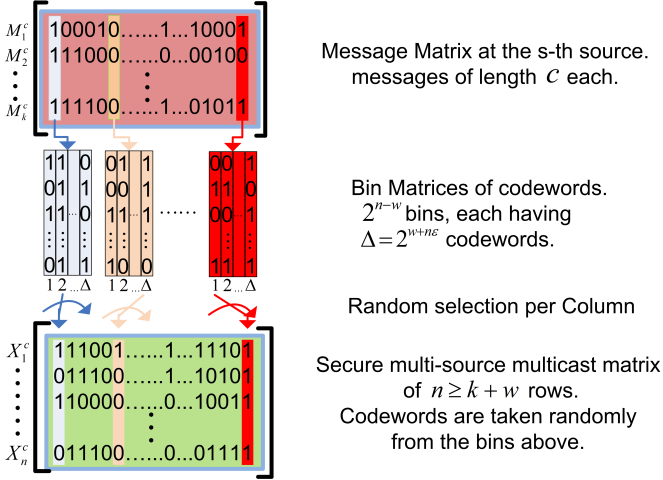


Figure 4: Binning and source encoding process for Strong-SMSM.

APPENDIX A CODE CONSTRUCTION AND A PROOF SKETCH FOR STRONG-SMSM

At each source node $s \in \{1, \dots, |S|\}$, we *randomly* map each column of the message matrix \mathbf{M}_s . As depicted in Figure 4, in the code construction phase, for each *possible column* of the s -th message matrix we generate a bin, containing several columns. The number of such columns *corresponds* to w , the number of packets that the eavesdropper can wiretap, in a relation that will be made formal in the sequel. Then, to encode, for each column of the message matrix, we randomly select a column from its corresponding bin. This way, a new, $n \times c$ message matrix \mathbf{X}_s is created. Specifically, a Strong-SMSM code at the s -th source node consists of a messages matrix \mathbf{M}_s of $\vec{M}_{s,1} \dots \vec{M}_{s,k}$ messages of length c bits over the binary field, we denote the set of matrices by \mathcal{M}_s ; A discrete memoryless source of randomness over the alphabet \mathcal{R} and some known statistics p_R ; An encoder,

$$f : \mathcal{M}_s \times \mathcal{R} \rightarrow \mathcal{X}_s \in \{0, 1\}^{n \times c}$$

which maps each message matrix \mathbf{M}_s to a matrix \mathbf{X}_s of codewords. This message matrix contains $n \geq k + w$ new messages of size c .

The need for a *stochastic encoder* is similar to most encoders ensuring information theoretic security, as randomness is required to confuse the eavesdropper about the actual information [26]. Hence, we define by R_k the random variable encompassing the randomness required for the k messages at the source node, and by Δ the number of columns in each bin. We may now turn to the detailed construction and analysis.

1) *Codebook Generation*: Set $\Delta = 2^{w+n\epsilon}$. Where $P(x) \sim \text{Bernoulli}(1/2)$, using a distribution $P(X^n) = \prod_{j=1}^n P(x_j)$, for each possible column in the message matrix generate Δ independent and identically distributed codewords $x^n(e)$, $1 \leq e \leq \Delta$. where $\epsilon \geq 1/n$.

2) *Source and legitimate Node encodings*: For each column i of the s -th message matrix \mathbf{M}_s , the s -th source node selects uniformly at random one codeword $x^n(e)$ from the i -th bin.

Therefore, the s -th source Strong-SMSM matrix \mathbf{X}_s contains c randomly selected codewords of length n , one for each column of the s -th message matrix. Then, the sources transmit linear combinations of the rows, with random coefficients. Nodes transmit random linear combinations of the vectors in \mathcal{S}_v , which is maintained by each node according to the messages received at the node.

The reliability in the Strong-SMSM algorithm is inherited from the reliability in RLNC. That is, if the min-cuts are $\rho(s, d) \geq k + w$ and $\rho(s, d) \geq (k + w)|S|$ for each $s \in S$ and $d \in D$, then $k + w = n$ messages can be transmitted reliably from each source to all destinations. Since the transformation \mathbf{M}_s to \mathbf{X}_s can be inverted, the destinations can decode the original messages.

A. Information Leakage at the Eavesdropper

We now prove the strong-security constraint is met, using techniques recently given in [25]. In particular, for the strong constraint, we wish to show that $I(\mathbf{M}_s; \mathbf{Z}_w)$ is small for all $s \in S$. We will do that by showing that given \mathbf{Z}_w , Eve's information, all possibilities for \mathbf{M}_s are equally likely, hence Eve has no intelligent estimation for \mathbf{M}_s .

Denote by \mathcal{C}_n the random codebook and by \mathbf{X}_s the set of codewords corresponding to $\vec{M}_{s,1} \dots \vec{M}_{s,k}$. To analyze the information leakage at the eavesdropper, note that Eve has access to at most w linear combinations on the rows of \mathbf{X}_s . We will assume these linear combinations are in fact independent, and since Eve has access to the coefficients, we will assume Eve can even use Gaussian elimination and have access to w rows from original matrix \mathbf{X}_s .

Next, note that the columns of \mathbf{X}_s are independent (by the construction of the codebook, creating \mathbf{X}_s is done independently per-column; c columns are used only to reduce the NC overhead). Hence, it suffices to consider the information leakage for each column $i \in \{1, \dots, c\}$ from \mathbf{X}_s separately. For each column i of \mathbf{M}_s , the encoder has Δ independent and identically distributed codewords, out of which one is selected. Hence, there is an exponential number of codewords, from the eavesdropper's perspective, that can generate a column in \mathbf{X}_s , and we require that Eve is still confused even given the w bit from each column.

Let $\mathbf{Z}_w(i)$ be the w bits Eve has from column i . Denote $l = n - w$. Define by $Sh(\mathbf{Z}_w(i), l)$ the set of all n -tuples consistent with $\mathbf{Z}_w(i)$, i.e.,

$$Sh(\mathbf{Z}_w(i), l) = \{b^n : b^n(\mathcal{S}_Z) = \mathbf{Z}_w(i)\},$$

where \mathcal{S}_Z denotes indices of the rows Eve has. Clearly, there are 2^l tuples in $Sh(\mathbf{Z}_w(i), l)$.

We assume Eve has the codebook, yet does not know which column from each bin is selected to be the codeword. Hence, we wish to show that given $\mathbf{Z}_w(i)$, Eve will have at least one candidate per bin. The probability for a codeword to fall in a given shell is

$$\begin{aligned} Pr(\mathbf{X}_s^n(i) \in \mathcal{C}_n \cap \mathbf{X}_s^n(i) \in Sh(\mathbf{Z}_w(i), l)) \\ = \frac{Vol(Sh(\mathbf{Z}_w(i), l))}{2^n} = \frac{2^{(n-w)}}{2^n}. \end{aligned}$$

In each bin of \mathcal{C}_n , we have $\Delta = 2^{w+n\epsilon}$ codewords. Thus, the average number of codewords Eve sees in her shell, *per bin* is

$$|\{m(i) : X^n(i) \in \mathcal{S}h(Z(i), l)\}| = \frac{2^{w+n\epsilon} * 2^{n-w}}{2^n} = 2^{n\epsilon}.$$

Hence, we can conclude that on average, and if $n\epsilon$ is not too small, for every column in \mathbf{M}_s Eve has a few possibilities *in each bin*, hence cannot locate the right bin. However, it is still important to show that all bins have (asymptotically) equally likely number of candidate codewords, hence Eve cannot locate a preferred bin.

To this end, we proved that the average number of codewords per column is very close to $2^{n\epsilon}$ with high probability. We wish to show that the probability that the actual number of options deviates from the average by more than ε is small. Define

$$\mathcal{E}_{C_1}(Z(i), l) := \Pr\{(1 - \varepsilon)2^{n\epsilon} \leq |m(i) : X_s^n(i) \in \mathcal{S}h(\mathbf{Z}_w(i), l)| \leq (1 + \varepsilon)2^{n\epsilon}\}.$$

By the Chernoff bound, we have

$$\Pr(\mathcal{E}_{C_1}(Z(i), l)) \geq 1 - 2^{-\varepsilon' 2^{n\epsilon}}.$$

Due to the super exponential decay in n , when taking a union bound over all columns, the probability that Eve decodes correctly some column is small. Hence, for Eve, all codewords are almost equiprobable and $I(\{\mathbf{M}_s\}_{s \in \mathcal{S}}; \mathbf{Z}_w) \rightarrow 0$.

REFERENCES

- [1] S.-Y. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE transactions on information theory*, vol. 49, no. 2, pp. 371–381, 2003.
- [2] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.
- [3] N. Cai and R. W. Yeung, "Secure network coding on a wiretap network," *Information Theory, IEEE Transactions on*, vol. 57, no. 1, pp. 424–435, 2011.
- [4] D. Silva and F. R. Kschischang, "Security for wiretap networks via rank-metric codes," in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*. IEEE, 2008, pp. 176–180.
- [5] S. El Rouayheb, E. Soljanin, and A. Sprintson, "Secure network coding for wiretap networks of type II," *Information Theory, IEEE Transactions on*, vol. 58, no. 3, pp. 1361–1371, 2012.
- [6] N. Cai and R. W. Yeung, "A security condition for multi-source linear network coding," in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*. IEEE, 2007, pp. 561–565.
- [7] Z. Zhang and R. W. Yeung, "A general security condition for multi-source linear network coding," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*. IEEE, 2009, pp. 1155–1158.
- [8] T. H. Chan and A. Grant, "Network coding capacity regions via entropy functions," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5347–5374, 2014.
- [9] N. Cai, "Valuable messages and random outputs of channels in linear network coding," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*. IEEE, 2009, pp. 413–417.
- [10] N. Cai and T. Chan, "Theory of secure network coding," *Proceedings of the IEEE*, vol. 99, no. 3, pp. 421–437, 2011.
- [11] D. Kobayashi, H. Yamamoto, and T. Ogawa, "Secure multiplex coding attaining channel capacity in wiretap channels," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8131–8143, 2013.
- [12] A. S. Mansour, R. F. Schaefer, and H. Boche, "Secrecy measures for broadcast channels with receiver side information: Joint vs individual," in *Information Theory Workshop (ITW), 2014 IEEE*. IEEE, 2014, pp. 426–430.
- [13] Y. Chen, O. O. Koyluoglu, and A. Sezgin, "On the individual secrecy rate region for the broadcast channel with an external eavesdropper," in *2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2015, pp. 1347–1351.
- [14] A. S. Mansour, R. F. Schaefer, and H. Boche, "The individual secrecy capacity of degraded multi-receiver wiretap broadcast channels," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 4181–4186.
- [15] —, "On the individual secrecy capacity regions of the general, degraded and gaussian multi-receiver wiretap broadcast channel," *IEEE Transactions on Information and Security*, 2016, vol. 11, no. 9, pp. 2107–2122, 2016.
- [16] M. Goldenbaum, R. F. Schaefer, and H. V. Poor, "The multiple-access channel with an external eavesdropper: Trusted vs. untrusted users," in *2015 49th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2015, pp. 564–568.
- [17] Y. Chen, O. O. Koyluoglu, and A. H. Vinck, "On secure communication over the multiple access channel," *International Symposium on Information Theory and Its Applications (ISITA), 2016 IEEE*, 2016.
- [18] A. S. Mansour, R. F. Schaefer, and H. Boche, "The individual secrecy capacity of the gaussian SISO and degraded gaussian MIMO multi-receiver wiretap channel," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2015, pp. 365–369.
- [19] L. Lima, M. Médard, and J. Barros, "Random linear network coding: A free cipher?" in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*. IEEE, 2007, pp. 546–550.
- [20] J. Claridge and I. Chatzigeorgiou, "Probability of partially solving random linear systems in network coding," *arXiv preprint arXiv:1607.04725*, 2016.
- [21] S. Deb, M. Médard, and C. Choute, "Algebraic gossip: A network coding approach to optimal multiple rumor mongering," *Information Theory, IEEE Transactions on*, vol. 52, no. 6, pp. 2486–2507, 2006.
- [22] A. Cohen, B. Haeupler, C. Avin, and M. Médard, "Network coding based information spreading in dynamic networks with correlated data," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 2, pp. 213–224, 2015.
- [23] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. M. Tolhuizen, "Polynomial time algorithms for multicast network code construction," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1973–1982, 2005.
- [24] J. Feldman, T. Malkin, R. A. Servedio, and C. Stein, "Secure network coding via filtered secret sharing." Citeseer.
- [25] B. K. Dey, S. Jaggi, and M. Langberg, "Sufficiently myopic adversaries are blind," in *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 1164–1168.
- [26] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.